AN EXPERIMENT IN THE USE OF "PERFECT" PROGNOSTIC CHARTS

PHILIP WILLIAMS, JR.

U.S. Weather Bureau, Salt Lake City, Utah

ABSTRACT

An experiment was conducted to determine how much if any, the 2-day forecasts for Utah, as issued by the forecast staff at Salt Lake City, would be improved if "perfect" prognostic charts were available for sea level, 700 mb., and 500 mb., for periods up to 48 hours. Results are discussed using the paired t test for differences between forecasts made from "perfect" prognostic charts and control forecasts, skill scores from precipitation contingency tables, and a non-parametric "Rank Method" comparison.

1. INTRODUCTION

There has been considerable discussion among forecasters from time to time concerning just how much weather forecasts would be improved if "perfect" prognostic charts of sea level pressure and frontal positions, as well as upper-air contours, were available. In the light of these discussions, it was decided to conduct an experiment among the forecasters at Salt Lake City to determine, insofar as possible, how much the 2-day Utah State forecasts would be improved if the prognostic charts referred to above, which are received via facsimile from the National Meteorological Center (NMC) in Suitland, Md. were "perfect."

2. PROCEDURE

This experiment was conducted in somewhat the same fashion as an experiment described by Panofsky and Brier [1] in their chapter on "Sampling Theory." In that experiment, forecasts were made first with incomplete observational data. After this initial forecast was completed, additional data were furnished to the forecaster and a revised forecast was prepared. Results were then analyzed to determine whether the additional data improved the forecasts, and if so, by how much. In the Salt Lake City experiment, forecasts were first made in the conventional manner, with most of the observational and chart material ordinarily available to the forecaster. Then perfect prognostic charts were supplied and a revised forecast made.

A month sufficiently removed from the time of the experiment (early 1959) had to be selected in order that forecasters would not recall the sequence of weather events. Also an active month with considerable precipitation and variation in temperature was needed, so as to make the test meaningful, yet with no outstanding or unusual weather events, for these might be retained in an attentive forecaster's memory. April 1953 met the requirements satisfactorily, and so was chosen for the experiment.

Forecasts were made from the 1730 MST surface map, for the three periods: tomorrow, tomorrow night, and the second day. Data available to the forecasters were the surface map, 700-mb. and 500-mb. charts, plotted raobs, and hourly sequence reports. Forecasts were made for five stations in Utah (Salt Lake City, Delta, Cedar City, Roosevelt, and Blanding) and included the 24-hour change in maximum or minimum temperature and precipitation for the three 12-hour periods. Forecasts were also prepared for tomorrow's maximum, tomorrow night's minimum, and the second day's maximum temperature for Salt Lake City only. Verification was tabulated on the present U.S. Weather Bureau "State forecast forms," using the rules for same, with a "trace" of precipitation considered as no precipitation. A temperature change of $\pm 9°$ F. was required to "break" a no-change forecast, while a change of $1°$ F. in the right direction verified a warmer or colder forecast. Since five stations were involved, 20 percent was deducted for each incorrect forecast. Thus scores for each period were either 0, 20, 40, 60, 80, or 100 percent for both precipitation and temperature change forecasts.

After the initial (control) forecasts were made, the perfect prognostic charts were made available to the forecaster, and he was instructed to make a revised forecast. The prognostic charts given to the forecaster were selected so as to represent fairly closely the pressure and contour prognostic material available by facsimile from NMC. At the time of the experiment, the latter provided sea level, 700-, and 500-mb. prognostic charts for approximately 12, 24, 36, and 48 hours in advance. Frontal intensities were labeled on the sea level prognostic charts, as is customarily done by NMC.

Five of the staff forecasters took part in the experiment, each making forecasts on from five to eight days, for the three periods mentioned above. Since forecasts were

made each day in April, there were 30 forecasts for comparison between perfect prognostic and control forecasts. No comparison was made between forecasters, as Panofsky and Brier did. Also, no forecaster made forecasts on consecutive days, as after completing his initial 48-hour forecast, the forecaster was given prognostic material for two days ahead. However, this restriction made the individual forecasts more independent, since most of the day-to-day serial correlation between overlapping forecasts was eliminated. It is well known that when forecasting under doubtful conditions a forecaster frequently tends to repeat his forecast for the overlapping period, applying a sort of "double or nothing" philosophy.

Since it was desired that full use be made of the prognostic charts, including the implicit temperature forecasting features [2], 1000-mb. charts were made from the sea level prognostic charts, and these were used to construct perfect 1000–700- and 1000–500-mb. thickness prognostic charts. Departures from normal thickness were also supplied to the forecaster. Each forecaster was aware of the relationship between mean virtual temperature of the layer and thickness; i.e., a thickness change of 200 ft. represents a 5.5° F. mean virtual change in the 1000–500-mb. layer, and an 11° F. change in the 1000–700-mb. layer. Using this information is, of course, merely extracting fully the information implicitly available from prognostic charts, perfect or imperfect [3, 4].

3. RESULTS

Average scores for perfect prognostic and control forecasts are shown in the top part of table 1. Results are given for precipitation and temperature forecasts for the three periods, both separately and combined. These periods are respectively 12–24, 24–36, and 36–48 hours in advance, and are labeled 1, 2, and 3 in the table. It can be seen (left and center columns) that the perfect prognostic scores (percent correct) for the State forecasts are better for all but the first-period precipitation, and are especially better for third-period precipitation. For precipitation forecasts, the longer the forecast interval, the more the perfect prognostics help, as perfect prognostic

scores essentially hold steady with increasing time, while control forecasts show a rapid decrease in accuracy.

The Weather Bureau State forecast temperature change verification scheme used here is relatively insensitive because there is a large overlapping where two different forecasts may be correct. It was felt that a forecast of actual temperature might give a more sensitive verification, so forecasters were also requested to predict actual temperatures for tomorrow's maximum, tomorrow night's minimum, and the second day's maximum at Salt Lake City. Average errors in degrees F. are shown in the right-hand columns of table 1. Since lower average errors represent better forecasts, differences are considered positive when the perfect prognostic forecasts are better.

It may be noted that both types of temperature forecasts are more improved for maxima (periods 1 and 3) than for minima (period 2) by use of the perfect prognostic charts. This is probably because maximum temperatures are more representative of the air mass than minimum temperatures in the area and season under consideration, and thus are more readily forecast from the perfect thickness prognostic charts. The correlation between maximum temperature and 1,000–700-mb. thickness (interpolated values) for April 1953 for four stations in Utah was 0.86, while between minimum temperatures and thickness it was only 0.66.

A number of statistical methods can be employed to test the significance of the results. First the t test, using the method of paired comparisons [5], was used to test whether perfect prognostic forecast scores were significantly different from control forecast scores. Those days for which the two forecast scores were identical were not included since these give no information about differences in forecasting skill. Results of this test are shown in the lower portion of table 1. It may be seen from the last row, labeled Probability, that the perfect prognostic third-period precipitation forecasts are better than the control forecasts at the 3 percent level of significance. No other precipitation scores are significantly different. For temperature change (State forecast) no individual period differences are statistically significant, but com-

TABLE 1.—*Comparative scores and results of* t *test*

	Precipitation (State forecasts)				Temperature change (State forecasts)				Salt Lake City actual temperature forecasts			
Period	1	2	3	Combined	1	2	3	Combined	1	2	3	Combined
N	30	30	30	30	30	30	30	30	30	30	30	30
Aver. {P. Progs	82.0	82.7	83.3	82.6	86.7	81.3	82.0	83.3	3.9	3.6	5.6	4.3
score {Control	86.0	79.3	72.0	79.0	78.7	78.0	74.3	77.0	5.3	4.6	7.3	5.7
Difference	−4.0	3.4	11.3	3.6	8.0	3.3	7.7	6.3	1.4	1.0	1.7	1.4

Results of *t* tests

	Precipitation (State forecasts)				Temperature change (State forecasts)				Salt Lake City actual temperature forecasts			
No. pairs	8	12	12	13	12	12	18	19	18	18	23	27
Mean diff	−15.0	8.3	31.7	9.4	20.0	8.3	12.8	10.7	2.2	1.7	2.2	1.5
Std. error	15.9	9.1	11.4	4.9	13.5	11.4	7.9	4.1	0.92	0.80	0.98	0.48
t	−0.94	0.92	2.78	1.93	1.48	0.73	1.61	2.61	2.49	2.15	2.27	3.14
Probability	0.38	0.38	0.03	0.11	0.17	0.48	0.13	0.02	0.02	0.05	0.03	<.01

TABLE 2.—*Precipitation contingency tables*

		Perfect prognostic forecasts												
		Period 1			Period 2			Period 3			All periods			
		Forecast			Forecast			Forecast			Forecast			
		R	NR	T	R	NR	T	R	NR	T	R	NR	T	
Observed	R	10	9	19	7	14	21	10	8	18	27	31	58	
	NR	18	113	131	12	117	129	17	115	132	47	345	392	
	T	28	122	150	19	131	150	27	123	150	74	376	450	
		Hits: 82 percent Skill score: 0.32			Hits: 83 percent Skill score: 0.25			Hits: 83 percent Skill score: 0.35			Hits: 83 percent Skill score: 0.31			

		Control forecasts												
Observed	R	12	7	19	3	18	21	2	15	17	17	40	57	
	NR	14	117	131	13	116	129	26	107	133	53	340	393	
	T	26	124	150	16	134	150	28	122	150	70	380	450	
		Hits: 86 percent Skill score: 0.45			Hits: 79 percent Skill score: 0.05			Hits: 73 percent Skill score: 0			Hits: 79 percent Skill score: 0.15			

bined perfect prognostic forecasts are significantly better at the 2 percent level. For Salt Lake City actual temperature forecasts in degrees F., the perfect prognostic forecasts are significantly better for all three periods, and the significance exceeds the 1 percent level for the combined forecasts.

Next, contingency tables were prepared for the precipitation forecasts for the three periods, both individually and combined, as shown in table 2. Chance skill scores and percent correct are also given. Since forecasts were made for five stations for each period each day, there are 150 individual rain or no-rain forecasts for each period for the entire month. Perfect prognostic forecasts show a lower skill score for the first period, but notably higher skill for the second and third periods.

A third comparison was made using the "Rank Method" as described by Panofsky and Brier [1]. In this method, no assumptions are made regarding the normality of frequency distributions. Forecast score differences are arranged in rank order (least difference, regardless of sign, is number 1, etc., after eliminating ties). The proper signs are then given to the ranks, and the rank differences for the fewest cases of the same sign are totalled. The smaller this number, the more significant the difference

between the two forecasts. Table 3 shows the results of this comparison. It may be seen that the perfect prognostic forecasts are significantly better at the 5 percent level for third-period precipitation and temperature change, and for combined temperature change forecasts. For Salt Lake City actual temperatures, the perfect prognostic forecasts are significantly better at the 5 percent level for all three periods, and at the 1 percent level for combined forecasts. Control precipitation forecasts are better than perfect prognostic forecasts for first-period precipitation, but not significantly so. This agrees with the results shown in tables 1 and 2.

4. INTERPRETATIONS AND CONCLUSIONS

Significant improvement, based on results as shown by the t test, is noted in the perfect prognostic forecasts for all periods for actual maximum and minimum temperature forecasts, for combined temperature change forecasts, and for third-period precipitation forecasts. All other differences are not significant at the 5 percent level, and for first-period precipitation, control forecasts were actually better than perfect prognostic forecasts. Also, perfect prognostic minimum temperature change forecasts showed

TABLE 3.—*Results of Rank Method comparison*

Period	Precipitation (State forecasts)				Temperature change (State forecasts)				Salt Lake City actual temperature forecasts			
	1	2	3	Combined	1	2	3	Combined	1	2	3	Combined
No. pairs	8	12	12	13	12	12	18	19	18	18	23	27
No. times perfect prog. better	2	7	9	9	7	7	12	13	12	12	16	18
No. times control better	6	5	3	4	5	5	6	6	6	6	7	9
Sum fewest rank diff	11	28	*12	19	23	30	*42	*41	*39	*41	*74	†68
5 percent level	4	15	15	18	15	15	42	49	42	42	75	124
1 percent level	0	7	7	10	7	7	31	35	31	31	56	96

*Significant at 5 percent level.
†Significant at 1 percent level.

very insignificant improvement. The Rank Method of comparison shows essentially the same results. All this would seem to indicate that there is a bigger and more difficult step in going from the perfect prognostic flow patterns and frontal positions to the actual weather forecast than was heretofore suspected. This is especially true if one considers the generally small fractional advance made from control forecast scores toward "perfection"; i.e. 100 percent for State forecasts, and zero error for actual temperature forecasts.

Although this is a small sample, first-period precipitation results indicate that short-period (less than 24 hours) prognostic charts of this type are of questionable value to the forecaster for precipitation forecasting under the conditions of this test. Apparently moisture, stability, vertical motion, and probably other types of prognostic charts are also needed in order to significantly improve short-period precipitation forecasts.

However, it does not seem plausible for the meteorologist to follow the exact statistical line of reasoning by rejecting any improvement that does not come up to the 5 percent level of significance, for even at the 10 percent or 20 percent level, it might be held that there is a fairly strong indication of effective improvement from the meteorological point of view.

Since this experiment was completed, the National Meteorological Center has begun transmitting forecasts of vertical motion, precipitation, temperature change, etc., which should be of considerable help to the forecaster.

ACKNOWLEDGMENTS

REFERENCES

1. H. A. Panofsky and G. W. Brier, *Some Applications of Statistics to Meteorology*, The Pennsylvania State University, 1958, 219 pp.
2. C. A. Kibler, C. M. Lennahan, and R. H. Martin, "Temperature Forecasting as an Implicit Feature in Prognostic Charts," *Monthly Weather Review*, vol. 83, No. 1, Jan. 1955, pp. 23–30.
3. C. A. Kibler and M. R. Rogers, "Anomalous Temperatures over the Rocky Mountain States, October 18–22, 1958," *Monthly Weather Review*, vol. 86, No. 10, Oct. 1958, pp. 416–422.
4. C. J. Boyden, "The Forecasting of Daily Mean Surface Temperatures from 1000–500 Millibar Thickness Lines," *The Meteorological Magazine*, vol. 87, No. 1030, Apr. 1958, pp. 98–105.
5. R. A. Fisher, *The Design of Experiments*, Second Edition, Oliver and Boyd, 1937, pp. 35–43.